

Computing Krippendorff's Alpha-Reliability

Krippendorff's alpha (α) is a reliability coefficient developed to measure the agreement between observers, coders, judges, raters, or measuring instruments. It emerged in content analysis but is widely applicable wherever two or more methods of processing data are applied to the same set of objects, units of analysis, or items and the question is how much they agree.

α 's general form is:

$$\alpha = 1 - \frac{D_o}{D_e}$$

where D_o is the observed disagreement:

$$D_o = \frac{1}{n} \sum_c \sum_k o_{ck \text{ metric}} \delta_{ck}^2$$

and D_e is the disagreement one would expect when the coding of units is attributable to chance rather than to the properties of these units:

$$D_e = \frac{1}{n(n-1)} \sum_c \sum_k n_c \cdot n_k \text{ metric} \delta_{ck}^2$$

The terms of the two disagreements will be defined and their computation exemplified below.

When observers agree perfectly, observed disagreement $D_o=0$ and $\alpha=1$, which indicates perfect reliability. When observers agree as if chance had produced the results, $D_o=D_e$ and $\alpha=0$, which indicates the absence of reliability. α would measure 0 if observers failed to observe and made up their data by throwing dice. When $\alpha=0$, data are totally uninformative of anything outside the process of generating them. Thus, for reliability considerations, α 's range is:

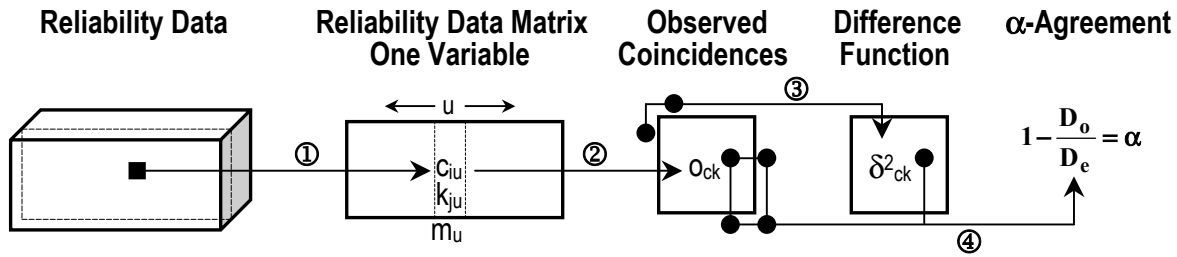
$$1 \geq \alpha \geq 0 \begin{cases} - \text{Systematic disagreement} \\ \pm \text{Sampling errors} \end{cases}$$

Unlike other specialized coefficients, α is a generalization of several reliability indices. It enables researchers to judge a variety of data with the same reliability standards. α applies to:

- Any number of observers, not just two
- Any number of categories, scale values, or measures
- Any metric or level of measurement (nominal, ordinal, interval, ratio, and more)
- Incomplete or missing data
- Large and small sample sizes alike, not requiring a minimum

α evaluates reliability one variable at a time. It offers other analytical possibilities not presented here.

Reliability data duplicate the process whose reliability is in question. Given such data, α -reliability can be computed in four computational steps, graphed below.



These four computational steps will be defined and demonstrated by four kinds of data of increasing generality:

- A. Binary or dichotomous data, two observers, no missing data
- B. Nominal data, two observers, no missing data
- C. Nominal data, multiple observers, missing data
- D. All metrics, multiple observers, missing data

A. Binary or dichotomous data, two observers, no missing data

① Construct a **reliability data matrix**; here, a 2 observer-by- r unit matrix, containing $2r$ values, c and k :

Units:	<i>1</i>	<i>2</i>	...	<i>u</i>	...	<i>r</i>	
Observers:	i:	c_{i1}	c_{i2}	...	c_{iu}	...	c_{ir}
	j:	c_{j1}	c_{j2}	...	c_{ju}	...	c_{jr}

For example, when two observers judge ten units, the 2-by-10 matrix contains 20 values:

Items judged:	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>	<i>10</i>
Meg:	0	1	0	0	0	0	0	0	1	0
Owen:	1	1	1	0	0	1	0	0	0	0

② Tabulate **coincidences within units**. Coincidence matrices account for the *values* contained in the reliability data matrix. They differ from the familiar contingency matrices, which account for *units*, not values. The importance of this difference becomes apparent in C. Into a coincidence matrix, units are entered twice, once as **c-k** pair and once as **k-c** pair. In the example, unit *1* is entered as a **0-1** pair of values *and* as a **1-0** pair of values. Unit *2* is entered as two **1-1** pairs of values, etc.:

Values:	0	1		0	1			
	0	o_{00}	o_{01}	n_0	0	o_{10}	o_{11}	n_1
	1	o_{10}	o_{11}	n_1	1	o_{01}	o_{00}	n_0
Number of Values:		n_0	n_1	n		14	6	20

Where o_{00} represents the ten **0-0** pairs within units *4, 5, 7, 8, and 10*. o_{01} represents the four **0-1** pairs in units *1, 3, 6, and 9*, and o_{10} represents the four **1-0** pairs in the same units. o_{11} represents the two **1-1** pairs found only in unit *2*. $n_0=14$ is the number of **0s** in the reliability data matrix, $n_1=6$ is the number of **1s**, and $n=2r=20$ is the total.

For A and only for these data, all mismatching coincidences are in two cells, $o_{01}=o_{10}$, here =4.

③ skip

④ Compute **α -reliability**:
$$\text{binary } \alpha = 1 - \frac{D_o}{D_e} = 1 - \frac{(n-1) \cdot o_{01}}{n_0 \cdot n_1}$$

In the example:

$$\text{binary } \alpha = 1 - \frac{(20-1) \cdot 4}{14 \cdot 6} = .095$$

B. Nominal data, two observers, no missing data

① Construct a **reliability data matrix** – just as in A above. For a 2-by-12 example:

Items judged:	1	2	3	4	5	6	7	8	9	10	11	12
Ben:	a	a	b	b	d	c	c	c	e	d	d	a
Gerry:	b	a	b	b	b	c	c	c	e	d	d	d

② Tabulate **coincidences within units**. Generically, and for the above example, the coincidence matrix is now:

Values:	1	.	k	.	.
1	o_{11}	o_{1k}	.	.	n_1
.
.
c	o_{c1}	o_{ck}	.	.	$n_c = \sum_k o_{ck}$
.
.
	n_1	n_k	.	.	$n \sum_c \sum_k o_{ck}$

	a	b	c	d	e
a	2	1	.	1	.
b	1	4	.	1	.
c	.	.	6	.	.
d	1	1	.	4	.
e	2
	4	6	6	6	2
	24				

Where $o_{ck} = \sum_u$ Number of **c-k** pairs in unit u specifically:

- $o_{ab} = 1$ **a-b** pair in unit 1
- $o_{ba} = 1$ **b-a** pair in unit 1
- $o_{aa} = 2$ **a-a** pairs in unit 2
- $o_{bb} = 4 = 2$ **b-b** pairs in unit 3
+ 2 **b-b** pairs in unit 4
and so forth.
- $n_a = 4$ is the number of **as**
- $n_b = 6$ is the number of **bs**
and so forth.
- $n = 24$ is the total number of values
in the reliability data matrix

③ skip

④ Compute **α -reliability**
$$\text{nominal } \alpha = \frac{(n-1) \sum_c o_{cc} - \sum_c n_c(n_c - 1)}{n(n-1) - \sum_c n_c(n_c - 1)}$$

In the example:

$$\text{nominal } \alpha = \frac{(24-1)(2+4+6+4+2) - (4(4-1) + 3[6(6-1)] + 2(2-1))}{24(24-1) - (4(4-1) + 3[6(6-1)] + 2(2-1))} = .692$$

C. Nominal data, multiple observers, missing data

① Construct a **reliability data matrix** – just as in A and in B above, but for m observers:

Units u:	1	2	.	.	.	u	r
Observers: 1	c ₁₁	c ₁₂	.	.	.	c _{1u}	c _{1r}
i	c _{i1}	c _{i2}	.	.	.	c _{iu}	c _{ir}
j	c _{j1}	c _{j2}	.	.	.	c _{ju}	c _{jr}
.
m	c _{m1}	c _{m2}	.	.	.	c _{mu}	c _{mr}

Number of values in u: m₁ m₂ . . . m_u . . . m_r

When data are missing, the numbers m_u of values in columns or units u are variable.

For a 4 observers-by-12 units example:

Units u:	1	2	3	4	5	6	7	8	9	10	11	12
Observer A:	1	2	3	3	2	1	4	1	2	.	.	.
Observer B:	1	2	3	3	2	2	4	1	2	5	.	.
Observer C:	.	3	3	3	2	3	4	2	2	5	1	3
Observer D:	1	2	3	3	2	4	4	1	2	5	1	.

Number m_u of values in unit u: 3 4 4 4 4 4 4 4 4 4 3 2 1 | 41

Note that 7 out of the 48 possible values in this matrix are missing. m_u varies from 1 to 4.

② Tabulate **coincidences within units**. The coincidence matrix appears as in B:

Values:	1	.	k	.	.	
1	o ₁₁	.	o _{1k}	.	.	n ₁
.
.
c	o _{c1}	.	o _{ck}	.	.	n _c = Σ _k o _{ck}
.
	n ₁	.	n _k	.	.	n = Σ _c Σ _k n _{ck}

	1	2	3	4	5	
1	7	4/3	1/3	1/3	.	9
2	4/3	10	4/3	1/3	.	13
3	1/3	4/3	8	1/3	.	10
4	1/3	1/3	1/3	4	.	5
5	3	3
	9	13	10	5	3	40

But, unlike in B:
$$o_{ck} = \sum_u \frac{\text{Number of c - k pairs in unit u}}{m_u - 1}$$

Note that each unit contains m_u(m_u-1) coincidences (pairs of values) that a coincidence matrix needs to account for. Unit **1** contains 3(3-1)=6 pairs of matching values **1**. It contributes 6/(3-1)=3 to the o₁₁ cell, one for each value. Unit **2** contains 4(4-1)=12 pairs, 6 matching **2-2** pairs, 3 mismatching **2-3** pairs, and 3 mismatching **3-2** pairs. It adds 6/(4-1)=2 to o₂₂, 3/(4-1)=1 to o₂₃, 1 to o₃₂, and 4 to the total n, thus fully accounting for its 4 values. Unit **6** contains 4(4-1)=12 pairs of mismatching values, each adds 1/(4-1)=1/3 to a different cell. The lone value **3** in unit **12** affords no comparisons and does not add to this account. Thus, the margins do not represent the 41 values that occur in the reliability data matrix, only the n=40 values that can be matched within units.

③ Skip

④ Compute **α -reliability** – just as in B

$$\text{nominal } \alpha = 1 - \frac{D_o}{D_e} = \frac{(n-1) \sum_c o_{cc} - \sum_c n_c(n_c - 1)}{n(n-1) - \sum_c n_c(n_c - 1)}$$

In the example: $\text{nominal } \alpha = \frac{(40-1)(7+10+8+4+3) - (9(9-1) + 13(13-1) + 10(10-1) + 5(5-1) + 3(3-1))}{40(40-1) - (9(9-1) + 13(13-1) + 10(10-1) + 5(5-1) + 3(3-1))} = .743$

D. Any metric, multiple observers, missing data

- ① Construct a **reliability data matrix** – just as in C
- ② Tabulate **coincidences** within units – just as in C
- ③ Insert the **difference function** $\text{metric } \delta_{ck}^2$ that is appropriate to the metric of the given data into the two disagreements D_o and D_e defined above.

Note that α accounts for different metrics or levels of measurement by weighing the observed (and expected) coincidences by the squared difference between the coinciding values. Differences can be expressed as mathematical functions and in the form of a table. The latter makes their relative magnitudes transparent. Interval and ratio metric differences are functions of the values being paired. Ordinal differences depend on their frequencies of using values. And nominal differences are added here to generalize step ④.

- **Nominal metric differences** – Two values either match, or they do not:

Nominal categories, names:

	a	b	c	d	e	f
a	0	1	1	1	1	1
b	1	0	1	1	1	1
c	1	1	0	1	1	1
d	1	1	1	0	1	1
e	1	1	1	1	0	1
f	1	1	1	1	1	0

$$\text{nominal } \delta_{ck}^2 = \begin{cases} 0 & \text{iff } c = k \\ 1 & \text{iff } c \neq k \end{cases}$$

- **Ordinal metric differences** – Values have the meaning of ranks and differences between ranks depend on how many ranks they are apart from each other. For example, with frequencies from data in C (and one unused rank added to show that it does not matter):

Ranks: 1st 2nd 3rd 4th 5th 6th

$$\text{ordinal } \delta_{ck}^2 = \left(\frac{n_c}{2} + \sum_{g < c} n_g + \frac{n_k}{2} \right)^2$$

where $c < k$

1 st	0	11 ²	22.5 ²	30 ²	32.5 ²	34 ²	$n_{1st} = 9$
2 nd	121	0	11.5 ²	19 ²	21.5 ²	23 ²	$n_{2nd} = 13$
3 rd	506	132	0	7.5 ²	10 ²	11.5 ²	$n_{3rd} = 10$
4 th	900	361	56	0	2.5 ²	4 ²	$n_{4th} = 5$
5 th	992	462	100	6.3	0	1.5 ²	$n_{5th} = 0$
6 th	1156	529	132	16	2.3	0	$n_{6th} = 3$

- **Interval metric differences** – Values differ algebraically:

Interval values:

$$\text{interval } \delta_{ck}^2 = (c-k)^2$$

	-1	0	1	2	3	4
-1	0	1 ²	2 ²	3 ²	4 ²	5 ²
0	1	0	1 ²	2 ²	3 ²	4 ²
1	4	1	0	1 ²	2 ²	3 ²
2	9	4	1	0	1 ²	2 ²
3	16	9	4	1	0	1 ²
4	25	16	9	4	1	0

- **Ratio metric differences** – Algebraic differences between two values are expressed relative to an absolute zero point. They are proportional to the magnitude of their values:

Ratio values:

$$\text{ratio } \delta_{ck}^2 = \left(\frac{c-k}{c+k} \right)^2$$

	0	1	2	3	4	5
0	0	$(\frac{1}{1})^2$	$(\frac{2}{2})^2$	$(\frac{3}{3})^2$	$(\frac{4}{4})^2$	$(\frac{5}{5})^2$
1	1	0	$(\frac{1}{3})^2$	$(\frac{2}{4})^2$	$(\frac{3}{5})^2$	$(\frac{4}{6})^2$
2	1	.11	0	$(\frac{1}{5})^2$	$(\frac{2}{6})^2$	$(\frac{3}{7})^2$
3	1	.25	.04	0	$(\frac{1}{7})^2$	$(\frac{2}{8})^2$
4	1	.36	.11	.02	0	$(\frac{1}{9})^2$
5	1	.44	.18	.06	.01	0

④ Compute **alpha-reliability**: $\text{metric } \alpha = 1 - \frac{D_o}{D_e} = 1 - (n-1) \frac{\sum_c \sum_{k>c} o_{ck \text{ metric}} \delta_{ck}^2}{\sum_c \sum_{k>c} n_c \cdot n_k \text{ metric } \delta_{ck}^2}$

In this computational formula, note that the sums enumerate only one of the two symmetrical off-diagonal triangles of the coincidence matrices o_{ck} .

Computations are demonstrated with the numerical data in C, interpreted as ordinal, interval and ratio data respectively. Zero frequencies are omitted from the lists of multiplications:

With data in C as **ordinal data**:
$$\text{ordinal } \alpha = 1 - (n-1) \frac{\sum_c \sum_{k>c} o_{ck \text{ ordinal}} \delta_{ck}^2}{\sum_c \sum_{k>c} n_c \cdot n_k \text{ ordinal } \delta_{ck}^2}$$

In the Example:

$$\text{ordinal } \alpha = 1 - (40-1) \frac{\frac{4}{3}1^2 + \frac{1}{3}225^2 + \frac{1}{3}30^2 + \frac{4}{3}115^2 + \frac{1}{3}19^2 + \frac{1}{3}7.5^2}{9 \cdot 13 \cdot 1^2 + 9 \cdot 10 \cdot 225^2 + 9 \cdot 5 \cdot 30^2 + 9 \cdot 3 \cdot 34^2 + 13 \cdot 10 \cdot 115^2 + 13 \cdot 5 \cdot 19^2 + 13 \cdot 3 \cdot 23^2 + 10 \cdot 5 \cdot 7.5^2 + 10 \cdot 3 \cdot 115^2 + 5 \cdot 3 \cdot 4^2} = .815$$

With data in C as **interval data**:
$$\text{interval } \alpha = 1 - (n-1) \frac{\sum_c \sum_{k>c} o_{ck \text{ interval}} \delta_{ck}^2}{\sum_c \sum_{k>c} n_c \cdot n_k \text{ interval } \delta_{ck}^2}$$

In the example:

$$\text{interval } \alpha = 1 - (40-1) \frac{\frac{4}{3}1^2 + \frac{1}{3}2^2 + \frac{1}{3}3^2 + \frac{4}{3}1^2 + \frac{1}{3}2^2 + \frac{1}{3}1^2}{9 \cdot 13 \cdot 1^2 + 9 \cdot 10 \cdot 2^2 + 9 \cdot 5 \cdot 3^2 + 9 \cdot 3 \cdot 4^2 + 13 \cdot 10 \cdot 1^2 + 13 \cdot 5 \cdot 2^2 + 13 \cdot 3 \cdot 3^2 + 10 \cdot 5 \cdot 1^2 + 10 \cdot 3 \cdot 2^2 + 5 \cdot 3 \cdot 1^2} = .849$$

With data in C as **ratio data**:
$$\text{ratio } \alpha = 1 - (n-1) \frac{\sum_c \sum_{k>c} o_{ck \text{ ratio}} \delta_{ck}^2}{\sum_c \sum_{k>c} n_c \cdot n_k \text{ ratio } \delta_{ck}^2}$$

In the example:

$$\text{ratio } \alpha = 1 - (40-1) \frac{\frac{4}{3} \frac{1^2}{3^2} + \frac{1}{3} \frac{2^2}{4^2} + \frac{1}{3} \frac{3^2}{5^2} + \frac{4}{3} \frac{1^2}{5^2} + \frac{1}{3} \frac{2^2}{6^2} + \frac{1}{3} \frac{1^2}{7^2}}{9 \cdot 13 \cdot \frac{1^2}{3^2} + 9 \cdot 10 \cdot \frac{2^2}{4^2} + 9 \cdot 5 \cdot \frac{3^2}{5^2} + 9 \cdot 3 \cdot \frac{4^2}{6^2} + 13 \cdot 10 \cdot \frac{1^2}{5^2} + 13 \cdot 5 \cdot \frac{2^2}{6^2} + 13 \cdot 3 \cdot \frac{3^2}{7^2} + 10 \cdot 5 \cdot \frac{1^2}{7^2} + 10 \cdot 3 \cdot \frac{2^2}{8^2} + 5 \cdot 3 \cdot \frac{1^2}{9^2}} = .797$$

Software for these and related computations is currently being developed.

Reference:

Klaus Krippendorff. *Content Analysis, an Introduction to Its Methodology*. Thousand Oaks, CA: Sage Publications, 1980. (New edition planned for 2003.)